情报学取向的"数据科学与大数据技术"专业人才培养*

陈沫 李广建 陈聪聪
北京大学信息管理系 北京 100871

摘要:[目的/意义]情报学作为计算性较强的人文社科学科,具备数据科学的特点,大数据相关技术能为情报学提供有效手段,因此,设立情报学取向的"数据科学与大数据技术"专业,培养具有大数据思维和扎实的情报学专业基础、掌握大数据处理技术及分析理论、并能将其应用于情报学理论与实践的复合型人才,能够促进情报学学科发展。[方法/过程]对国内外大数据相关专业的培养目标和课程设置模式进行调研,总结学科教育发展特色,结合国内的教学资源情况,设计情报学取向的大数据专业人才培养计划。[结果/结论]情报学与大数据既有联系又有不同,情报学为大数据提供了基础研究方法,促进了多源数据融合,拓宽了大数据的应用领域;大数据为情报学研究提供了更多的可能性、丰富了研究方法,提升了研究效率、扩大了研究范围。情报学取向的大数据与数据科学专业人才的培养,既有其必要性,又有其不可替代性。学科体系需进一步改革创新,设立更多紧密结合大数据与情报学专业的新兴课程,如文献大数据分析、用户大数据研究等,系统地培养学生将情报学领域知识与计算机技术和大数据技术融合创新的能力。

关键词:情报学教育 人才培养 大数据 数据科学
分类号:G250
DOI:10.13266/j.issn.0252-3116.2019.12.001

1 引言

随着大数据时代的到来,社会各个行业的数据呈现出高速增长的状态,数据资源已经成为了重要的现代战略资源。数据科学学科的建设及大数据人才的培养因此受到了学术界与教育界的广泛关注。自北卡罗莱纳州立大学2007年首次设立"数据分析"硕士专业以来,美国已经有多所高校陆续开设"数据分析"或"数据科学"等相关专业,不仅涉及到传统的计算机、数学和统计学等院系,而且还涉及到图书情报等人文社科类的院系,例如,阿姆斯特丹大学的人文、档案和信息研究院设立了"信息研究:数据科学"专业。与国外相比,我国高校在大数据与数据科学教育开展相对较晚,最早是北京航空航天大学于2013年开设了"数据科学"硕士专业,可以说开我国正式的大数据高等教育的先河。2016年,教育部批准首次北京大学、对外经济贸易大学、中南大学三所高校增设"数据科学与大数据技术"本科专业。截至2018年底,国内一共有283所高校获批设立该本科专业。在这些高校中,承担该专业建设的院系各有不同,不同的学科院系和课程内容反映了一个学校及学院的办学特色和人才特色,以及考虑培养出来的学生是否受到了社会的欢迎。近年来,情报学领域对大数据的研究越来越重视,对具有图书情报特色的大数据人才培养得到了本学科关注,学界多有探讨。苏新宁指出,大数据为情报学的发展带来了机缘,情报学教学体系的变革要扬长避短,专注情报技术的研究[1]。巴志超等人认为,情报学要深入推进各门类学科与情报学之间多方式、大跨度的广泛交叉,从而建设多元开放、互动协同发展的学科生态群[2]。苏日娜等人以15所iSchools高校为调研对象,讨论了图书馆与情报学学科下的数据科学定位、数据科学与传统图书情报课程结合和人才培养等问题[3]。

在此背景下,本文探讨了数据科学与大数据专业本科教育的相关问题。首先对国内外大数据专业的培养目标、教学内容和课程计划进行了广泛调研,在此基础上总结了大数据专业教学内容的特点;随后结合国

* 本文系国家社会科学基金重点项目"大数据环境下的计算型情报分析方法与技术研究"(项目编号:14ATQ005)研究成果之一。
作者简介:陈沫(ORCID:0000-0003-3452-0664),博士研究生;李广建(ORCID:0000-0002-2897-6246),系主任,教授,通讯作者,E-mail:ligj@pku.edu.cn;陈聪聪,硕士研究生。
收稿日期:2018-11-15 修回日期:2019-02-06 本文起止页码:5-11 本文责任编辑:杜杏叶

内的实际情况，总结了建设基于情报学的数据科学与大数据技术专业的必要性；最后，初步设计了情报学取向的大数据专业人才培养课程设置和培养目标。

2 国内外大数据教育的调研

早在 20 世纪末，国外就有高校尝试开设数据科学相关的专业及课程。发展至今，国外大数据教育的学科领域呈现出多样化的特点，计算机科学、数学与统计、经济与管理、生物与医疗、科学与技术、新闻与传媒等多个学科领域均有开设大数据专业，课程设置及研究方向各具特色，阮敬等[4]总结出国外大数据相关硕士人才培养主要有七个方向，包括：信息系统、数据科学、健康医疗、商业分析、应用统计、商务智能和 MBA 大数据方向。在美国，由 60 多所学校的信息学院组成的 iSchool 联盟院校中，已经有多所图书情报院系开设了数据科学与大数据技术的专业[5]。

与国外大数据相关的专业大多面向硕士研究生不同，我国高教界更关注本科教育。2016 年，教育部颁布的《2015 年度普通高等学校本科专业备案和审批结果》[6]中，首次增设了“数据科学与大数据技术”本科专业，并批准了北京大学、对外经济贸易大学、中南大学三所高校设立该专业，此后分别于 2017 年度[7]和 2018 年度[8]分别新增了 32 所高校与 248 所高校设立这一专业，见图 1。截至 2018 年，国内一共有 283 所高校获批“数据科学与大数据技术”专业，其中，985 及 211 高校共 36 所，占比 13%。该专业学制为四年，授予工学学位或理学学位，其中，北京大学、人民大学和上海财经大学等高校在多个院系开设了这一专业。但目前国内尚未有图书情报或其他人文社科院系开设该专业。

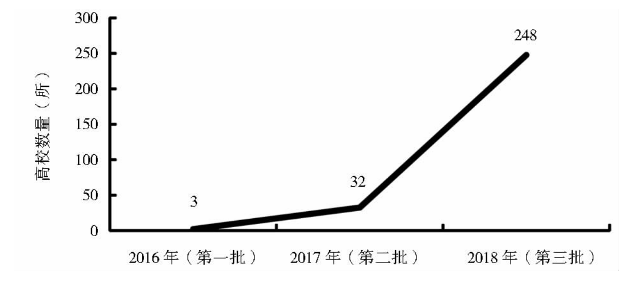


图 1 全国开设“数据科学与大数据技术”专业本科的高校数量

本文对国内外大数据相关专业的培养目标、教学内容和课程计划进行了调研，发现大数据专业教育有以下两个显著特点，一是已经形成了层次性的课程体系，二是形成了与所依托专业相结合的特色化侧重方向。

2.1 层次性的课程体系

数据科学是一门以大数据为研究对象的交叉性学科，其理论基础主要来自于计算机科学、应用数学以及信息管理与信息系统等相关学科，同时，由于大数据本身只存在于各个不同领域和不同的行业之中，不是独立于任何行业或学科领域的简单数据堆积，因此，处理大数据还需要具备大数据所属领域或行业的专门知识，例如经济学、法学、生物学、医学等学科知识。在实践上，国内外大数据专业已经在课程体系上形成了比较清晰的三个层次，见表 1。其中，最顶层是应用层，是利用大数据及数据科学的理念和方法去解决问题的应用场景，实际上就是诸如生物医学、企业管理、科技管理、电子商务等各类产生大数据的行业所属的学科领域，前常见的包括这些学科专业的核心和关键知识，具体体现是这些学科的核心课程；中间层是方法层，由大数据及数据科学的理念、方法和技术构成，一方面，这一层次的课程为学生提供处理大数据的基本理念、方法、技术与工具。另一方面，与应用层的知识相结合，提供有关特定领域大数据处理的模型和算法，训练学生对应用数据进行理解、抽象和建模，从而在底层的计算平台上予以处理和分析；最底层是基础层，主要包括统计学、数学、计算理论与技术等基础学科课程，这是目前公认的大数据与数据科学的基础领域学科。

表 1 层次性的课程体系

课程所属体系层次	层次介绍
基础层	数学、统计学、计算机理论与技术等基础课程
方法层	涉及大数据方法与技术的方法课程
应用层	面向应用领域的相关课程

例如，对外经济贸易大学信息学院的“数据科学与大数据”专业的课程设置[9]比较明显地体现出了层次性的特征，其课程体系划分为数学基础、计算机基础、大数据的技术与方法、经管法四大课程模块，分别对应着基础层、方法层、应用层三大层次，具体课程体系见表 2。数学与计算机基础课程模块主要培养学生传统信息技术行业的基本能力，也为数据科学及大数据方法与技术相关的方法课程提供基础知识；大数据技术与方法相关课程旨在让学生掌握数据分析与数据建模方法、大数据处理技术，从而具备理解、抽象、分析、建模、开发等一系列解决数据科学问题的能力；经管法相关课程模块主要涉及经济学、管理学、法学相关的课程，能够让学生对特定数据行业的知识有所了解，支持学生从事经济、金融、管理、物流、商务等领域的数据分析工作的素养与知识储备。

表 2　对外经济贸易大学信息学院“数据科学与大数据”专业课程体系

课程所属体系层次	课程分类	课程名称
基础层	数学基础	数学分析、线性代数、概率论与数理统计、随机过程
	计算机基础	C + + 程序设计、计算机组成原理、数据结构、操作系统原理、JAVA 程序设计、数据库系统原理、计算机网络、Android 程序开发基础、Web 应用程序设计
方法层	数据科学	多元统计分析、回归分析与计量、时间序列分析、器学习与数据挖掘、数据可视化原理及应用、大数据存储与管理、Spark 大数据处理、Storm 实时大数据处理、大数据分析实践
应用层	经管法	微观经济学、货币银行学、管理学原理、金融风险管理、电子商务、推荐系统原理、电子金融、网络营销、文本挖掘与自然语言理解

2.2　特色化的侧重方向

大数据专业教育的第二个特点是已经形成了各具特色化的侧重方向。目前,各个国内外高校的大数据专业都是在已有专业的基础上展开的,通常都是与原有专业研究的领域与方向相结合,在人才培养目标上具有一定的侧重点,其培养模式基本上可以归纳为以下三类。

2.2.1　以技术为特色的大数据教育　以技术为特色的大数据教育是目前国内外大数据及数据科学相关专业主流的侧重方向,主要是培养偏向于大数据技术研究与应用的人才,培养学生设计新的大数据相关系统、技术或程序的计算机开发能力,更加强调学生掌握大数据技术前沿的状态及趋势,具备大数据平台与工具的开发、数据处理系统的构建等实践能力。这类专业的课程体系与计算机技术与理论结合得更加紧密,通常设立在计算机学院或信息学院,例如华威大学(University of Warwick)计算机科学学院的数据分析(data analytics)专业、电子科技大学计算机科学与工程学院下的数据科学与大数据技术专业、北京邮电大学计算机学院下的数据科学与大数据技术专业等,表 3 是以北京邮电大学为例列出的以技术为特色的课程设置。

北京邮电大学的数据科学与大数据技术专业的课程体系中包含了大量的计算机科学与技术学科的基础课程,以培养学生多元化的大数据应用编程能力。

2.2.2　以计算为特色的大数据教育　以计算为特色的大数据教育则更偏向于培养大数据算法与通用建模的人才,培养学生开发新的大数据相关算法、模型或方法的能力,要求学生精通数据分析理论与方法,能够解决各类较大规模或结构复杂数据的处理问题。这类专业的课程体系中往往包含大量的数学与统计学课程,在计算与建模方面的学习程度更深,通常设立在数学

表 3　北京邮电大学计算机学院数据科学与大数据技术专业课程体系

课程分类	课程名称
数学与自然科学基础	数学分析、高等数学、大学物理、物理实验、线性代数、概率论与随机过程、组合数学、运筹学、数学建模与模拟、矩阵理论与方法、离散数学、网络科学
计算机和数据科学与大数据技术	计算导论与程序设计、电路与电子学基础、数字逻辑与数字系统、形式语言与自动机、数据结构、算法设计与分析、操作系统、编译原理与技术、计算机组成原理、计算机系统结构、计算机网络、数据库系统原理、大数据技术基础、NoSQL 数据库技术、数据可视化、数据仓库与数据挖掘、机器学习
综合应用	数据科学导论、软件工程、计算导论与程序设计、面向对象分析与设计(Java)、并行计算与 GPU 编程、数据采集与管理、数据分析与计算、数据服务与应用、技术拓展、网络科学、人文社科类

学院或统计学院,例如爱丁堡大学(University of Edinburgh)的数据科学的运筹研究专业和纽约大学(New York University)的库朗数学学院开设的数据科学硕士专业[10],下面以纽约大学为例对其具体课程设置进行列表,如表 4 所示:

表 4　纽约大学数学学院数据科学硕士专业课程体系

课程性质	课程名称
必修课程	Introduction to Data Science(数据科学导论)
	Probability and Statistics for Data Science(数据科学中的概率统计)
	Machine Learning(机器学习)
	Big Data(大数据)
	Capstone Project and Presentation(Capstone 项目与演示)
选修课程	Inference and Representation(推理与表示)
	Deep Learning(深度学习)
	Natural Language Processing with Representation Learning(自然语言处理与表征学习)
	Natural Language Understanding and Computational Semantics(自然语言理解与计算语义)
	Optimization-based Data Analysis(基于优化的数据分析)
	Optimization and Computational Linear Algebra(最优化和计算线性代数)

纽约大学是全球首个创建“数据科学”(data science)硕士专业的高校,专业招生面向具有数学、计算机科学和应用统计学背景的学生,侧重于开发数据科学的新方法。与本科课程设置相比,面向硕士研究生的课程安排较少,除了基础的数据科学概论、大数据和机器学习等课程外,该专业的课程设置有更多的计算与算法类课程,如基于优化的数据分析、优化与计算线性代数、数据科学领域的概率论与统计学等课程。

2.2.3　以专业领域为特色的大数据教育　以专业领域为特色的大数据教育通常培养的是特定行业领域的大数据专业人才,往往要求学生在掌握大数据技术与

方法的同时,着重对特定领域相关的专业知识具有足够的了解,例如金融、生物、医学等专业领域。这种类型的专业教育对学生的大数据技术或建模能力要求较低,只需要其能够在特定领域使用相关方法与技术即可。目前,这类专业通常设立在特定领域相应的学科学院下,例如伦敦国王学院(King's College London)在数字人文学院开设的社会文化大数据专业[11],其将大数据链接到文化、法律与伦理、地理、公共卫生和社会生活,旨在从艺术和人文学的角度来解决大数据的理论和实践问题,具体课程设置如表5所示:

表5 伦敦国王学院数字人文学院社会文化大数据专业

课程性质	课程名称
必修课程	Theorising Big Data(大数据理论)
	Big Data in Practice: Co-laboratories, Tools and Methods(实践中的大数据:联合实验室、工具和方法)
	From Data to Insight: Cultural and Social Analytics(从数据到洞察力:文化和社会分析)
	Big Data and the Law: Foundations, Regimes and Principles, Challenges(大数据与法律:基础、制度与原则、挑战)
选修课程	Digital Arts and Culture(数字艺术与文化)
	Editorial models for Digital Texts: Theory & Practice(数字文本编辑模型:理论与实践)
	Web Technologies(网络技术)
	Digital Publishing(数字出版)
	Communication & Consumption of Cultural Heritage(文化遗产的传播与消费)
	Material Culture of the Book(书籍的物质文化)
	Applied Visualization for Cultural Heritage(文化遗产应用可视化)
	Open Culture(开放文化)
	Ontologies of Digital Media(数字媒体的本体论)
	Social Media: Protest & Political Campaigning(社交媒体:抗议和政治活动)
	Metadata & Systems for Digital Assets & Media(用于数字资产和媒体的元数据和系统)
	Management for Digital Content Industries(数字内容产业的管理)
	Digital Media, Digital Marketing(数字媒体,数字营销)
	Curating & Preserving Digital Culture(组织与保存数字文化)
	Crowds & Clouds - Digital Ecosystems(人群和云-数字生态系统)

在我国,以专业领域为特色的数据科学与大数据技术专业大多与管理和经济金融领域结合,例如复旦大学的管理学院、北京大学元培学院和北京大学数学科学学院[12]等,下面以北京大学数学科学学院为例进行课程设置列表,见表6。

北京大学是首批设立"数据科学与大数据技术"专业的三所高校之一,目前由信息科学技术学院、元培学院和数学科学学院分别开设该专业。其中数学科学学院的该专业为互联网金融分析方向,课程体系中除

表6 北京大学数学科学学院数据科学与大数据专业课程体系

课程分类	课程名称
数学与统计学基础	实变函数、微分几何、偏微分方程、拓扑学、微分流形、数论、群与表示、基础代数几何、密码学、数理统计、随机过程、多元统计分析、测度论、统计计算、抽样调查、时间序列分析、非参数统计、试验设计、随机分析、生存分析、贝叶斯理论与算法、统计学习、生物信息学、数值代数、数值分析最优化方法
计算机基础	程序设计与技术、计算机科学基础、计算机图形学、算法设计与分析、软件工程、操作系统、数据库、数据结构与算法、信息安全
数据科学	大数据分析中的算法、人工智能、数据整合与清理、自然语言处理、数据仓库与数据挖掘、数据库、数据结构与算法、信息安全、机器学习
经济金融学科	金融数学引论、寿险精算、奉献力量、衍生证券基础、金融经济学、金融数据分析

了大量的数学与统计学基础课程,还涉及到经济金融学科的应用课程,培养学生将大数据理论与方法技术应用于经济金融领域的能力。综上所述,与国外的数据科学相关专业环境相比,国内的"数据科学与大数据技术"等相关专业教育尚在起步阶段,存在诸多不足之处。国外的相关专业发展已久,而国内的本科教育刚正式进行了三年左右,硕士教育也多处于探索阶段,整个学科教育环境尚不成熟。在国外,该学科领域呈现出多样化的特点,广泛涉及除计算机外的多个学科领域,而国内的该专业多设立在计算机技术或经济与管理相关的学院,另有少数设立在数学、医疗等相关学院,其他领域鲜有涉猎,发展方向尚未实现多样化。此外,虽然国内大多数院校的该专业课程体系层次化较为明显,但在应用层面上,多是简单保留原有专业的特色课程,将大数据与本专业紧密结合的课程设置仍然较少,课程设置的创新性不足。

虽然国内的"数据科学与大数据技术"专业开始较晚、覆盖的专业领域多样化不足,且课程设置的专业化不足,但在国内学科设立的初期,批准设立该专业的本科院校覆盖面较广,其中985及211高校仅占约13%,有利于大数据相关人才的全面培养。

3 基于情报学的数据科学与大数据专业教育

3.1 学科建设意义

情报学(Information Science)是指以现代信息技术和软科学的研究方法作为主要手段,以信息的采集、选择、评价、分析和综合等系列化的加工为基本过程,形成新的或增值的情报产品,来满足用户不同的特定需求,为不同层次科学决策服务的人文社科类学科[13]。近年来,随着大数据技术的发展,图书情报领域的人才

市场需求也逐渐提出要掌握数据科学和大数据技术相关技能,以 2018 年高校应届毕业生秋季招聘为例,从应届生论坛展示的部分数据来看,全国逾百个与情报学相关的岗位招聘中明确提到希望应聘者掌握大数据相关技能或有相关经历,如广东数鼎科技有限公司在招聘情报信息分析人员的信息中要求有数据分析能力,熟练使用 SAS 等数据统计软件[14];长沙市图书馆在招聘运维专员时也强调应聘者要有较强的大数据分析能力及良好的定性分析能力[15]。

设立基于情报学的大数据专业教育,属于上述提到的第三种以专业领域为特色的大数据教育模式,这种类型的专业教育重点要解决以下两个问题:首先是相关的专业是否需要大数据教育,或者说现有的专业是否存在着与大数据必然的联系,是否存在着相关联的基础;第二个问题是基于所建立的大数据教育与已有的专业教育之间的关系,即两者是否存在差别。以下结合情报学本身的特点,试图对以上两个问题做出回答。

3.1.1 情报学与大数据的关联性 大数据分析(Big Data Analytics,BDA)是大数据理念与方法的核心,是指对海量＝类型多样、增长快速且内容真实的数据(即大数据)进行分析,从中找出可以帮助决策的隐藏模式、未知的相关关系以及其他有用信息的过程[16]。与前文中对情报学的定义相比,两者都是以信息和数据作为基础资源和研究对象。数据最早来源于测量,所谓“有根据的数字”,是对客观世界测量结果的记录,不是随意产生的。而现如今,数据的内容开始扩大,它不仅指有根据的数字,还统称一切保存在电脑中的信息,而这些都不再是测量得到的数据,而是对客观世界的一种记录[17]。大数据不仅包括结构化的数据,也包括非结构化的数据,如传统情报学所研究的文献、文本以及行为等数据都可以成为一种大数据,情报学与大数据技术都要对数据进行采集和加工处理,进而使用数据带来的价值。例如,Min. Song 等通过对大数据量级的生物医学领域文献数据进行引文分析,得出了药物与疾病之间的隐含联系,为生物医学研究的临床实验做出贡献[18]。冯元为[19]提出通过构建人物关联关系的知识图谱,帮助查询人物的生活轨迹、社会关系等基本信息,从而为公安情报工作带来便利。

所谓“大数据”都是由具体的数据构成的,不存在抽象的“大数据”。不管是计算机领域研究的“大数据”还是统计学、数学领域研究的“大数据”,都是与其研究领域相结合,解决对应领域的特定问题,而不是单纯的“大数据”问题。大数据价值的挖掘依赖于领域知识和经验,大数据价值的发挥,是通过领域学科来扩散和发挥的,图书情报机构也是大数据行业中的一员,与大数据息息相关。例如,医院或医疗研究机构通过将大数据技术应用到医疗领域,使用医疗大数据来帮助提升医疗服务;商业咨询公司、银行等机构通过分析商业大数据来帮助进行评级和决策;交通和公共管理部门通过处理城市交通大数据来帮助车辆调度,减轻交通压力。图书馆、中国知网、SCI 等图书情报机构也通过文献大数据分析为大众提供图书和文献信息服务。

情报学最早出现于第二次世界大战之后,经过长期的发展积累,已形成了一套独具特色的研究方法与理论体系。而大数据的概念产生于 20 世纪末期,在其发展过程中,也融合了多种学科的理论与方法,其中就包括情报学。情报学领域的诸多方法可以丰富大数据方法－体系,帮助提升大数据技术效率。例如情报学中的引文分析方法可帮助筛选重点信息源,从而提升文本分析和挖掘的效率,成为大数据挖掘的前期辅助处理。

3.1.2 情报学取向的数据科学与大数据专业的不可替代性 情报学与其他领域的大数据分析既有联系又有区别,两者的研究对象虽然都是数据,但情报学相关的大数据主要研究文献大数据、文本大数据和行为大数据。情报学更加注重资源的收集与整理服务,在进行数据分析时有明确的任务导向和既定的模式,而其他学科导向的大数据分析更加注重对数据本身的分析,旨在挖掘新的模式。

情报学也积累了一套处理这类数据的理论和方法,包括分类与主题的思想和方法、规范档、元数据、文献计量、引文分析、情报分析的理论和方法等。情报学与大数据理念方法相结合,将丰富情报学的内容,提高情报工作的精度和效率,例如,以往多是通过以元数据为代表的文献外部特征进行分析,例如关键词、摘要、主题词等,而依托于数据科学与大数据技术,能够实现对文献内部更细粒度的研究,例如词向量分析、文本挖掘等,从而开启了文献领域研究的新视野,丰富了研究的内容。

此外,随着情报学学科和社会的发展,情报分析呈现出新的特点,逐渐着重计算化、多源数据融合[20]分析和深入的内容分析,需要基于大量数据构建大规模的知识图谱、社会网络等,而这些都是传统的情报学所不能完成的,需要有新的理念和方法加以支撑,这就需要情报学与大数据技术相结合,既能加深情报学的研究深度、提升研究效率,又能促进大数据方法和技术的发展,实现两学科的共同进步。

因此,情报学取向的数据科学与大数据专业与其

他学科取向的该专业之间不能相互取代,都有其不同学科特定的研究内容与侧重,都在做其他学科领域无法完成的研究。另外,情报学取向的大数据教育也不等同于情报学教育的扩展和扩充,不能被情报学教育取代。情报学取向的数据科学与大数据专业的设置旨在填补当前学科专业教育的空白,既能丰富大数据与数据科学专业的学科应用,又能顺应时代特色,促进图书情报学科的大数据与数据科学专业人才的培养。

3.2 培养目标

情报学取向的"数据科学与大数据技术专业"旨在培养具有大数据思维和扎实的情报学专业基础、掌握大数据处理技术及分析理论、熟悉大数据技术与方法,能够运用统计向量分析、机器学习、文本挖掘等技术,从大量数据中提取对情报学研究与实践有意义的信息,并以社会网络和知识图谱等技术展现的高层次复合型人才。系统地培养学生掌握大数据应用中解决情报学中各种典型问题的解决办法,实际提升学生解决实际问题的能力,培养学生将情报学领域知识与计算机技术和大数据技术融合、创新的能力。对符合专业培养和毕业要求的学生授予管理学学士学位。

3.3 课程设置

结合以上四个模块,基于"基础层－方法层－应用层"的层次体系,对情报学取向的数据科学与大数据技术专业课程体系设置如图 2 所示:

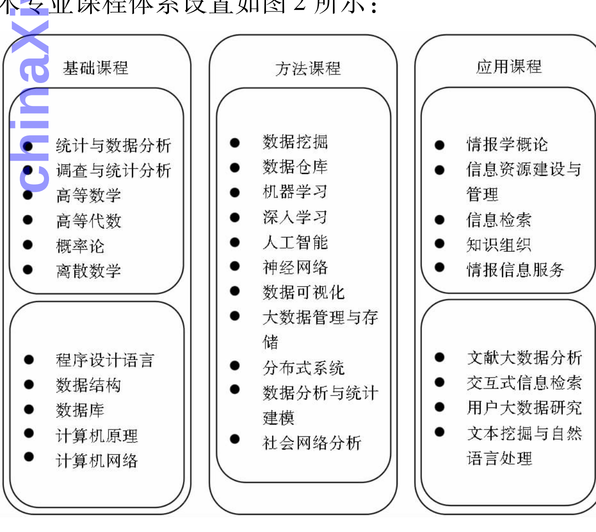


图 2 情报学取向的"数据科学与大数据技术专业"课程体系

其中,底层是基础层课程,代表统计学、数学与计算机应用技术等基础课程;中间层则是方法层,代表数据科学方法相关课程,对大数据进行理解、抽象、建模,然后在底层的计算平台上予以实现的相关理论和技术课程;最顶层是应用层,代表需要利用数据科学去解决情报学问题的应用场景,课程包括情报学基础课程和情报学与大数据融合的前沿特色课程。

4 结论

面对大数据时代大量的数据科学类多层次的人才需求,数据科学与大数据技术专业不仅需要计算机、数学和统计学学科背景,同样需要情报学等以数据为基础的人文社科领域共同促进发展。情报学与大数据既有联系又有不同,情报学与大数据技术的研究对象都包括数据,而大数据价值的发挥要依靠各个学科的交叉融合,同时图书情报机构也是大数据行业中的一员。情报学为大数据提供了基础研究方法,促进了多源数据融合,拓宽了大数据的应用领域;大数据为情报学研究提供了更多的可能性、丰富了研究方法,提升了研究效率、扩大了研究范围。情报学取向的大数据与数据科学专业人才的培养,既有其必要性,又有其不可替代性,将填补该专业教育在中国的空白。情报学取向的大数据与数据科学专业既不是传统情报学的简单扩充,也不能被其他学科取向的大数据与数据科学专业所取代。

国内的大数据相关专业教育尚处于初步探索阶段,发展时间较短,依然有很多不足之处,覆盖的学科领域需要应市场与社会发展的需求而不断丰富,情报学等图书情报领域学科的加入也可以促进大数据技术的发展、学科体系的丰富和不同学科的交叉融合,帮助培养更多的跨学科应用与科研人才。此外,学科体系也需进一步改革创新,促进各院校设立更多紧密结合大数据与原专业的新兴课程,进而丰富课程体系、提升教学质量与效果。

参考文献:

[1] 苏新宁. 大数据时代情报学学科崛起之思考[J]. 情报学报, 2018,37(5):451－459.

[2] 巴志超,李纲,周利琴,等. 数据科学及其对情报学变革的影响[J]. 情报学报,2018,37(7):653－667.

[3] 苏日娜,杨沁. LIS 学科中数据科学课程体系设置研究——以 iSchools 高校课程调研为中心[EB/OL]. [2019－03－19]. http://kns. cnki. net/kcms/detail/44. 1306. G2. 20181009. 0828. 002. html.

[4] 阮敬,刘宏晶,纪宏. 国外大数据硕士人才培养的经验与启示——基于大数据文本挖掘[J]. 统计与信息论坛,2017,32(9):29－36.

[5] 曾粤亮. 国外 iSchools 数据科学项目人才培养模式的特点与启示[J]. 知识、学习与管理,2018,184(4):109－118.

[6] 中华人民共和国教育部. 教育部关于公布 2015 年度普通高等学校本科专业备案和审批结果的通知[EB/OL]. [2018－07－01]. http://www.moe.edu.cn/srcsite/A08/moe_1034/s4930/201603/t20160304_231794.html.
[7] 中华人民共和国教育部. 教育部关于公布 2016 年度普通高等学校本科专业备案和审批结果的通知[EB/OL]. [2018－07－01]. http://www.moe.gov.cn/srcsite/A08/moe_1034/s4930/201703/t20170317_299960.html.
[8] 中华人民共和国教育部. 教育部关于公布 2017 年度普通高等学校本科专业备案和审批结果的通知[EB/OL]. [2018－07－01]. http://www.moe.gov.cn/srcsite/A08/moe_1034/s4930/201803/t20180321_330874.html.
[9] 对外经济贸易大学. 信息学院数据科学与大数据技术专业培养方案[EB/OL]. [2018－07－01]. http://jwc.uibe.edu.cn:82/ADMINUI/UploadFiles/files/1_linjingtan/201710201540083994.pdf.
[10] New York university. MS in data science curriculum[EB/OL]. [2018－07－01]. https://cds.nyu.edu/academics/ms-in-data-science/ms-curriculum.
[11] 北京大学. 北京大学本科教学计划(2016 版)(理科卷)[EB/OL]. [2018－07－01]. http://www.dean.pku.edu.cn/userfiles/upload/download/201804242014397998.pdf.
[12] King's College London. Big data in culture & society[EB/OL]. [2018－07－01]. https://www.kcl.ac.uk/study/postgraduate/taught-courses/digital-humanities-ma.aspx.
[13] 包昌火. 情报研究方法论[M]. 北京:科学技术文献出版社,1990.
[14] 广东数鼎科技有限公司. 招聘信息[EB/OL]. [2019－02－02]. http://www.yingjiesheng.com/job-003-813-711.html.
[15] 湖南省长沙市图书馆. 招聘信息[EB/OL]. [2019－02－02]. http://www.yingjiesheng.com/job-003-850-173.html.
[16] Whitehouse. Big data across the federal government[EB/OL]. [2018－07－01]. http://www.Whitehouse.gov/sites/default/files/microsites/ostp/big_data_fact_sheet_final.pdf.
[17] 涂子沛. 大数据及其成因[J]. 科学与社会,2014,4(1):14－26.
[18] SONG M, KANG K, JU Y. Investigating drug-disease interactions in drug-symptom-disease triples via citation relations[J]. Journal of the Association for Information Science and Technology,2018,69(11):1355－1368.
[19] 冯元为. 基于知识图谱构建人物关系的设计与实现[D]. 重庆:重庆大学,2016.
[20] 李广建,化柏林. 大数据分析与情报分析关系辨析[J]. 中国图书馆学报,2014,40(5):14－22.

作者贡献说明：
陈沫：负责数据采集、调研分析及文章撰写与修改；
李广建：提供研究的主题、研究思路以及文章审阅与修改；
陈聪聪：负责国内高校专业的调研分析及文章撰写。

Talent Training of Information Science Oriented “Data Science and Big Data Technology” Discipline

Chen Mo　Li Guangjian　Chen Congcong
Department of Information Management, Peking University, Beijing 100871

Abstract: [**Purpose/significance**] As a highly computational humanities and social sciences discipline, information science has the characteristics of data science. Big data related technology can provide an effective means for information science. Therefore, establishing the information science oriented “Data Science and Big Data Technology” discipline that aims at training inter-disciplinary talents can promote the development of information science, and these talents have big data thinking and solid information science professional basis, mastering big data processing technology and analytical theory applied to the theory and practice of information science. [**Method/process**] We investigated the training objectives and curriculum model of big data related majors at home and abroad, summarized the development characteristics of subject education, and combined the teaching resources of domestic to design a training plan for big data professionals with information science orientation. [**Result/conclusion**] There are both connections and differences between information science and big data. Information science provides basic research methods for big data, promotes the fusion of multiple source data, and broadens the application fields of big data. Big data provides more possibilities for information science research, enriches research methods, improves research efficiency, and expands the scope of research. There are both necessity and irreplaceability of the talent training of information science oriented “data science and big data technology” discipline. It will fill the gap in this professional education in China. It is necessary to further reform and innovate the disciplinary system, and to establish more emerging courses that closely integrate big data and information science, such as literature big data analysis and user big data research, so as to systematically cultivate students' ability to integrate knowledge in the field of information science with computer technology and big data technology.

Keywords: information science education　talent training　big data　data science